

## CLAIMS

What is claimed is:

1. A system that facilitates classifying messages in connection with spam prevention, comprising:
  - a component that receives a set of the messages;
  - a first classification component that identifies a subset of the messages as SPAM or flagged for further analysis, and temporarily delays further classification of the subset of messages; and
  - a second classification component that after the delay period classifies the subset of messages.
2. The system of claim 1, the second classification component identifying some of the subset of messages as good based on a lack of sufficient new negative information.
3. The system of claim 1, the second classification component identifying some of the subset of messages as good based on new positive information other than a close match to a good message.
4. The system of claim 1, the messages are classified as spam or flagged or delayed based on a lack of information.
5. The system of claim 1, the messages are reclassified based on updated information from a machine learning spam filter.
6. The system of claim 2, wherein the lack of sufficient new negative information comprises the lack of appearance of similar messages in honeypots.
7. The system of claim 2, wherein the lack of sufficient new negative information comprises a lack of user complaints on similar information.

8. The system of claim 2, the lack of sufficient new negative information comprises information from polling users about at least a subset of messages.
9. The system of claim 2, the lack of sufficient new negative information comprises a low volume of similar messages.
10. The system of claim 8, the messages are classified as similar based on the sender's identity.
11. The system of claim 10, the sender's identity is classified based on his IP address.
12. The system of claim 8, the similarity of messages is based on the URLs contained in the messages.
13. The system of claim 1, messages initially classified as spam are deleted based on new information.
14. The system of claim 1, the spam is permanently deleted.
15. The system of claim 1, the spam is moved to a deleted messages folder.
16. The system of claim 1, further comprising a feedback component that receives information relating to the first and/or second classification component(s)', and employs the information in connection with training a spam filter or populating a spam list.
17. The system of claim 1, wherein the messages comprise at least one of: electronic mail (e-mail) and messages.

18. The system of claim 1, wherein the component that receives a set of the messages is any one of an e-mail server, a message server, and client e-mail software.
19. A server employing the system of claim 1.
20. An e-mail architecture employing the system of claim 1.
21. A computer readable medium having stored thereon the components of claim 1.
22. The system of claim 1, further comprising a quarantine component that quarantines the subset of messages based at least in part upon identification as flagged for further analysis by the first classification component.
23. The system of claim 1, the quarantining effected via placing the subset of messages in a folder separate from other messages.
24. The system of claim 1, the folder is visible or invisible to a user.
25. The system of claim 1, further comprising an identification component that identifies a source associated with a high occurrence of the subset of messages.
26. The system of claim 1, further comprising a time-stamp component that stamps at least one of an original arrival date on the message and a release date when classification of the message resumes.
27. The system of claim 1, the subset of messages excludes at least one of messages from senders on safelists, messages readily identified and classified as spam, messages readily identified and classified as good.

28. The system of 1, the first classification component determines length of delay before classification of the subset of messages is performed.

29. The system of claim 28, the length of delay is based at least in part upon at least one of the following:

- amount of time until a next scheduled filter update;
- amount of time until download of new or updated filter; and
- spam probability score assigned to respective messages in the subset.

30. A method for classifying messages, comprising:  
receiving a set of messages to classify; and  
based on lack of sufficient information, temporarily delaying classification on at least a subset of the messages as either spam or good or initially classifying the subset of messages as untrustworthy or suspicious.

31. The method of claim 30, further comprising a machine learning filter trained to determine the likelihood of quarantining aiding a correct eventual classification.

32. The method of claim 30, further comprising resuming classification when at least one of the following occurs:

- a quarantine period elapses; and
- additional information about the subset of messages has been obtained to facilitate classification of the respective messages in the subset as either spam or good.

33. The method of claim 30, the subset of messages excluding messages that is readily classified as spam or good or is from senders on a safelist.

34. The method of claim 30, temporarily delaying classification of the message when based at least in part upon at least one of the following:

- sender's IP address on the message has not been seen before;

sender's domain has not been seen before;  
sender's domain lacks a reverse IP address;  
a forward lookup on the sender's domain does not at least approximately  
match the sender's IP address;

the message comprises at least one of an embedded domain name, an  
embedded macro, and an executable file;

the message comprises conflicting evidence of good and spam messages;

the message originates from a location associated with spam;

the message is written in a language associated with spam;

the message comprises primarily an image; and

the message comprises HTML.

35. The method of claim 30, further comprising delivering at least a subset of  
suspicious messages

36. The method of claim 35, the subset of suspicious messages is delivered to  
their respective intended recipients and their actions facilitate determining whether the  
subset of messages is spam or good.

37. The method of claim 35, the subset of messages for which feedback is  
sought is a fixed percentage of messages or a fixed quantity of messages per sender that  
are temporarily delayed from classification.

38. The method of claim 35, the subset of messages for which feedback is  
sought is allowed to get through without classification as either spam or good to facilitate  
learning more about the messages.

39. An API that facilitates classifying messages by quarantining comprising:  
calculating a spam probability score for incoming messages;  
quarantining at least a subset of messages based at least in part upon their  
respective spam probability scores; and

recommending a quarantine time.

40. The API of claim 39, further comprising quarantining at least a subset of messages until the next filter download, at which time the filter determines whether to continue quarantining or resume classification of the messages; and repeating until a final classification of either spam or good is made.

41. The API of claim 39, further comprising communicating between server and client that server filter(s) has quarantined the respective message for a time period; and reducing a client filter quarantine time.

42. A system for classifying messages, comprising:  
means for receiving a set of messages to classify; and  
means for based on lack of sufficient information, temporarily delaying classification on the message as either spam or good or initially classifying the message as untrustworthy or suspicious.